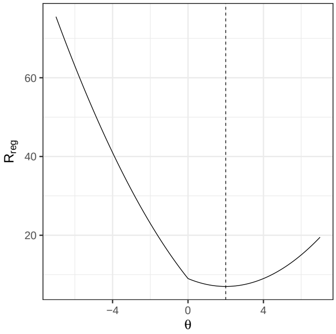
Introduction to Machine Learning

Soft-thresholding and L1 regularization (Deep-Dive)



Learning goals

- Understand the relationship between soft-thresholding and L1 regularization

SOFT-THRESHOLDING AND L1 REGULARIZATION

In the lecture, we wanted to solve

$$\min_{\boldsymbol{\theta}} \tilde{\mathcal{R}}_{\text{reg}}(\boldsymbol{\theta}) = \min_{\boldsymbol{\theta}} \mathcal{R}_{\text{emp}}(\hat{\theta}) + \sum_j \left[\frac{1}{2} H_{j,j}(\theta_j - \hat{\theta}_j)^2\right] + \sum_j \lambda|\theta_j|$$

with $H_{j,j} \geq 0, \lambda > 0$. Note that we can separate the dimensions, i.e.,

$$\tilde{\mathcal{R}}_{\text{reg}}(\boldsymbol{\theta}) = \sum_j z_j(\theta_j) \text{ with } z_j(\theta_j) = \frac{1}{2} H_{j,j}(\theta_j - \hat{\theta}_j)^2 + \lambda|\theta_j|.$$

Hence, we can minimize each z_j separately to find the global minimum.

If $H_{j,j} = 0$, then z_j is clearly minimized by $\hat{\theta}_{\text{Lasso},j} = 0$. Otherwise, z_j is strictly convex since $\frac{1}{2} H_{j,j}(\theta_j - \hat{\theta}_j)^2$ is strictly convex and the sum of a strictly convex function and a convex function is strictly convex.

SOFT-THRESHOLDING AND L1 REGULARIZATION

For convex functions, every stationary point is a minimum. Thus, we analyze the stationary points $\hat{\theta}_{\text{Lasso},j}$ of z_j for $H_{j,j} > 0$.

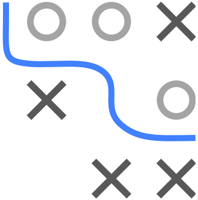
For this, we assume we already know the sign of the minimizer and then derive conditions for which our assumption holds.

So, first we consider the cases $\hat{\theta}_{\text{Lasso},j} > 0, \hat{\theta}_{\text{Lasso},j} < 0$.

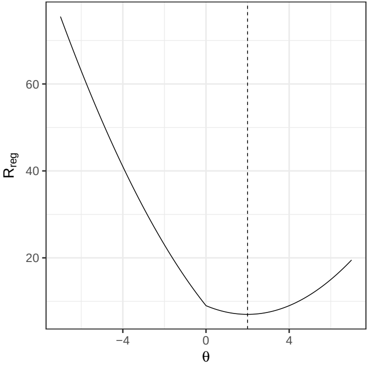
NB:

- For $\theta_j > 0 : \frac{d}{d\theta_j}|\theta_j| = \frac{d}{d\theta_j}\theta_j = 1$.
- For $\theta_j < 0 : \frac{d}{d\theta_j}|\theta_j| = \frac{d}{d\theta_j}\left(-\theta_j\right) = -1$.

SOFT-THRESHOLDING AND L1 REGULARIZATION



1) $\hat{\theta}_{\text{Lasso},j} > 0$:



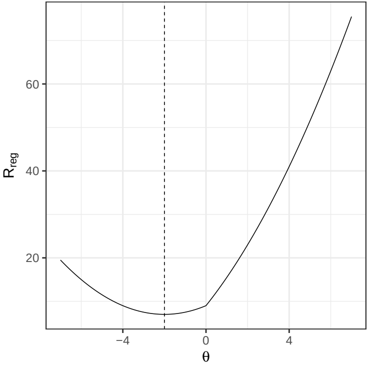
$$\frac{d}{d\theta_j} z_j(\theta_j) = H_{j,j}\theta_j - H_{j,j}\hat{\theta}_j + \lambda \overset{!}{=} 0$$

$$\Rightarrow \hat{\theta}_{\text{Lasso},j} = \hat{\theta}_j - \frac{\lambda}{H_{j,j}} > 0$$

$$\iff \hat{\theta}_j > \frac{\lambda}{H_{j,j}}$$

SOFT-THRESHOLDING AND L1 REGULARIZATION

2) $\hat{\theta}_{\text{Lasso},j} < 0$:

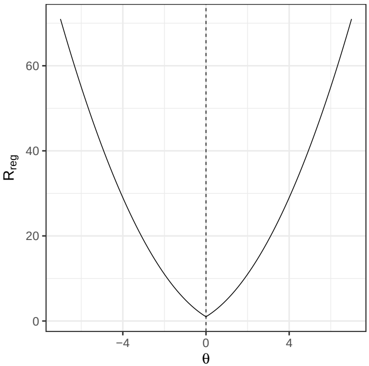


$$\frac{d}{d\theta_j} z_j(\theta_j) = H_{j,j}\theta_j - H_{j,j}\hat{\theta}_j - \lambda \overset{!}{=} 0$$

$$\Rightarrow \hat{\theta}_{\text{Lasso},j} = \hat{\theta}_j + \frac{\lambda}{H_{j,j}} < 0$$

$$\iff \hat{\theta}_j < -\frac{\lambda}{H_{j,j}}$$

SOFT-THRESHOLDING AND L1 REGULARIZATION



$\Rightarrow$ If $\hat{\theta}_j \in [-\frac{\lambda}{H_{j,j}}, \frac{\lambda}{H_{j,j}}]$ then z_j has no stationary point with

$$\hat{\theta}_{\text{Lasso},j} < 0 \text{ or } \hat{\theta}_{\text{Lasso},j} > 0.$$

However, a unique stationary point must exist since z_j is strictly convex for $H_{j,j} > 0$. This means, here, z_j is strictly monotonically decreasing (increasing) for $\theta_j < 0$ ($\theta_j > 0$).

$$\Rightarrow \hat{\theta}_{\text{Lasso},j} = \begin{cases} \hat{\theta}_j + \frac{\lambda}{H_{j,j}} & \text{, if } \hat{\theta}_j < -\frac{\lambda}{H_{j,j}} \text{ and } H_{j,j} > 0 \\ 0 & \text{, if } \hat{\theta}_j \in [-\frac{\lambda}{H_{j,j}}, \frac{\lambda}{H_{j,j}}] \text{ or } H_{j,j} = 0 \\ \hat{\theta}_j - \frac{\lambda}{H_{j,j}} & \text{, if } \hat{\theta}_j > \frac{\lambda}{H_{j,j}} \text{ and } H_{j,j} > 0 \end{cases}$$